Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 2 of 14

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 1 of 14

Module Objectives

- Describe National Longitudinal Study of 1972 (NLS-72), High School and Beyond (HS&B), and National Education Longitudinal Study of 1988 (NELS:88) weights that must be applied to ensure that estimates are representative of the specific study population
    - Describe the types of weights within NLS-72, HS&B, ad NELS:88 and provide guidance regarding how to select the appropriate weight for specific analyses
- Describe appropriate procedures for calculating standard errors

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 3 of 14

NLS-72, HS&B, and NELS:88 Sampling Designs and Weights

- There are not any short cuts when working with Secondary Longitudinal Studies from 1972-2000
- Even though there are **major** design features common across NLS-72, HS&B, and NELS:88, each of these data sets contain unique variations in methodologies used that require analysts to become familiar with the technical documentation and reports that have been written
- There is not always consistency in how the studies were designed, how the data files were set up, and how the weights were created
- **In order to fully understand the sampling designs and associated weights, you MUST review each study's documentation and understand how the data was collected, coded, and weighted to ensure accurate analyses**

Slide 4 of 14

NLS-72 Sample and Study Design

Stratified two-stage probability sample

- Stage 1: School selection
  - All public and private schools in the 50 states and the District of Columbia that enrolled 12th graders during the 1971-72 school year
    - Excluded students from schools for the physically or mentally handicapped, legally confined students, and those enrolled in other high schools (e.g., vocational schools)
  - Oversampled schools in low-income areas and schools with a high proportion of minority group enrollment
  - Final sample: 1,061 high schools provided base year data (257 schools were added during the first follow-up)

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 5 of 14

NLS-72 Sample and Study Design (Continued)

Stratified two-stage probability sample

- Stage 2: Student selection
  - Goal was 18 seniors per school
  - Final sample: 16,683 students (4,450 students were added during the first follow-up)

Slide 6 of 14

HS&B Sample and Study Design

Stratified two-stage probability sample

- Stage 1: School selection
  - Public and private high schools in the 50 states and the District of Columbia
  - Oversampled public schools with high percentages of Hispanic students, Catholic schools with high percentages of minority students, alternative public schools, and private schools with high-achieving students
  - Final sample: 1,015 high schools provided base year data

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 7 of 14

HS&B Sample and Study Design (Continued)

Stratified two-stage probability sample

- Stage 2: Student selection
  - 36 sophomores and 36 seniors were sampled from each school
  - Final sample: (30,030 sophomores and 28,240 seniors)

Slide 8 of 14

NELS:88 Sample and Study Design

Stratified two-stage probability sample

- Stage 1: School selection
  - 40,000 public and private schools serving approximately 3,000,000 8th graders in the 50 states and the District of Columbia were identified
  - Oversampled Asian/Pacific Islanders and Hispanic students and private schools
  - Final sample: 1,052 high schools provided base year data

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 9 of 14

NELS:88 Sample and Study Design (Continued)

Stratified two-stage probability sample

- Stage 2: Student selection
  - Randomly selected 24 to 26 students per school
  - Oversampled Asian/Pacific Islander, Hispanic, and private school students
  - Excluded approximately 5% of students on rosters
    - Mental disabilities (~3%)
    - Physical disabilities (<1%)
    - Language difficulties (~2%)
  - Final sample: 24,599 students

Slide 10 of 14

Use of Weights – Review

- Used to make estimates from sample data representative of target population
- Account for differential selection probabilities and differential patterns of response/nonresponse for the NLS-72, HS&B, and NELS:88
- Longitudinal studies like NLS-72, HS&B, and NELS:88 that have multiple components across multiple rounds of data collection have several possible weights for analysis of data within and across rounds

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 11 of 14

Working with Weights from NLS-72, HS&B, and NELS:88

Even though the studies are similar, it is important to **read the documentation** regarding sampling and weights for the specific data set of interest

- NLS-72 – when working with all six rounds of data
  - One sets of weights must be used when working with data from BY through F4 data (22,652 cases)
  - Another set of weights must be used when working with F5 data (14,489 cases)
- HS&B sophomore cohort – there are two BY weights available for analysis
  - 30,030 sample members as of 1980
  - 14,825 sample members that were part of the BY and F5
  - Use the BY through F5 weight BYWT
  - It is important to note that for HS&B, reduced data sets are provided in the current data releases (containing 14,825 sophomores and 11,995 seniors)

Slide 12 of 14

Working with Weights from NLS-72, HS&B, and NELS:88 (Continued)

- Even though the studies are similar, it is important to **read the documentation** regarding sampling and weights for the specific data set of interest
  - NELS:88 – there is only one BY cross-sectional weight BYQWT
    - BYQWT is only provided on the NELS:88/92 data file; it is not included on the NELS:88/94 or NELS:88/2000 data files
- The only way to learn how to appropriately use weights within the Secondary Longitudinal Studies from 1972-2000 is to read the documentation

Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 13 of 14

Calculating Standard Errors using NLS-72, HS&B, and NELS:88 Data

- In the common module titled, 'Statistical Analysis of NCES Datasets Employing a Complex Sample Design', two standard error calculation procedures were discussed: Replication Techniques and Taylor Series linearization
  - Replication is a method that calculates appropriate standard errors based on differences between estimates from the full sample and a series of created subsamples (replicates)
  - The Taylor Series linearization method uses PSU and strata identifiers to compute the appropriate standard errors
- Analysts using NLS-72, HS&B, and NELS:88 data will need to consult the technical documentation associated with each study to determine which method should be used to compute standard errors

Slide 14 of 14

Module Summary and Resources

Summary

- Described NLS-72, HS&B, and NELS:88 weights that must be applied to ensure that estimates are representative of the specific study population
  - Described the types of weights within NLS-72, HS&B, and NELS:88 and provided guidance regarding how to select the appropriate weight for specific analyses
- Described appropriate procedures for calculating standard errors

Resources

- Weights that must be applied
- Calculating appropriate standard errors
- Appropriate use of weights
- NLS-72
- HS&B
- NELS:88